# IMAGE DENOISING VIA ENERGY ORIENTED SPARSE REPRESENTATION[1]

**Jungyu Kang[2,3], Hoyong Jang[2,4], Chang D. Yoo[2,5]**

**[2]Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea**
**[3]cmiller2air@kaist.ac.kr, [4]hoyong.jang@kaist.ac.kr, [5]cdyoo@kaist.ac.kr**

This paper considers a denoising of zero-mean white additive Gaussian noise from an image. Sparse representation technic has shown good performance in denoising noises while keeping the image from being blurred. To optimize the conventional sparse representation algorithm, a concept of energy function is combined. An energy function to specify the noise level, estimation technic to find the optimal level of representation, result of the experiments, and further possible improvements are introduced.

## Introduction

This paper considers a traditional image denoising problem. Here the noise is considered to be additive noise. Here our goal is to find estimation of original image x from given noised image y. Although usually the noise n varies through conditions such as camera, environment, etc., here we only consider zero-mean white Gaussian with standard variance σ. The Gaussian distribution becomes a general model for various noises since the sum of random variable tend to a Gaussian distribution.

Since this is a very traditional problem and studied for a long time, various approaches were conducted. Although the results varies with the texture or complexity of the input image, algorithms such as BM3D [4], SKR[10], and K-LLD [3] showed best performances in denoising zero-mean white Gaussian noises. Yet, it is considered that it can still be improved. Among many competitive denoising approaches, our method focuses on using sparse representation[6].

More specifically, our method aim to improve existing denoising algorithm using sparse representation [6]. This paper aims to optimize the reconstruction rate of the sparse representation using a specifically designed energy function instead of simple sum of square error.

Fig 1 – Denoising result

The remainder of the paper is structured as follows: Section 2 briefly provides preliminary knowledge about dictionary learning, sparse representation algorithm, and the conventional sparse representation based denoising algorithm by Elad et al.. Section 3 introduces the proposed algorithm about designing the energy function and estimating the original image. Section 4 provides example results and evaluations of the proposed algorithm. Finally, section 5 concludes the paper with descriptions of future works.

## Related Works

Dictionary learning is a data representing algorithm by solving a matrix factorization problem. Let us consider a set $Y = \{ y^{1}, \dots, y^{n} \} \in R^{m \times n}$ which is a set of m dimensional n signals. In our problem, the **Y** will be the input noised image and **y** will be

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.NRF-2011-0017202 and NRF-2010-0028680)

the patches of the noised image. The goal of dictionary learning is to reconstruct the input signal **Y** as linear combination of two matrices, **D** and **A**. Here $\mathbf{D} \in R^{m \times k}$ denotes a m-dimensional trained dictionary containing k basis vectors. $\mathbf{A} \in R^{k \times n}$ denotes a decomposition coefficient for n signals. In detail, the dictionary learning aims to build an appropriate over-completed dictionary that can represent the input signals well with given constraints. This is achieved by optimizing the following equation.

$$\min_{\boldsymbol{D} \in D, A \in A} \left|\left|Y - DA\right|\right|_2^2 \quad s.t \sum_{i=1}^{n} \boldsymbol{\Omega}(\alpha_i) \quad (1)$$

Here, the constraint omega is a constraint to prevent overfitting and exploit certain property such as nonnegativity, sparsity, or structure to the coefficients. To optimize the equation, various optimization technics has been proposed [5, 7, 8]. In this paper, K-SVD algorithm [1] is used to learn **D** and **A** in a same time.

After learning the appropriate dictionary using the KSVD algorithm, we aim to reconstruct the input signals. However, since we want to enforce sparsity to our solution, l1 norm is used as the constraint term. Therefore the equation above can be reformulated as below:

$$\min_{A \in A} \sum_{i=1}^{n} \left|\left|\boldsymbol{y}_i - \boldsymbol{D}\boldsymbol{\alpha}_i\right|\right|_2^2 + |\boldsymbol{\alpha}_i| \quad (2)$$

Here, since the above equation is a convex function, the optimal solution can be obtained with convex optimization [2]. However, the convex optimization usually takes long time to get a solution, so we use OMP [9] instead.

The Orthogonal Matching Pursuit (OMP) algorithms is a greedy algorithm which seeks the best basis vector from the given basis vectors D to reconstruct the input signal in each iteration. It projects the whole basis vectors to the signal and finds the basis vector which has the maximum value. The corresponding coefficient value to the basis is obtained by simply performing a pseudo-inverse, which provides a same solution with the least square problem. The residual is updated by subtracting the projection of the chosen basis. These procedures are iteratively done until the result reaches the stopping rule.

The major advantages of this algorithm are that it works very fast, it is easy to implement, and the stopping rule can be designed as the user wants. Usually the number non-zero terms or reconstruction error are used as the stopping rule. However, in this paper, we introduce the energy function to control the sparsity and reconstruction error of the solution.

## Proposed Algorithm

To start, we first introduce the conventional image denoising algorithm using sparse representation [6]. In the algorithm, a dictionary for the sparse representation is prepared from either natural images or the noised image itself using the K-SVD algorithm. Using these dictionaries, the algorithm defines an optimization problem below

$$arg\min_{\boldsymbol{X}, \alpha_{ij}} \sum_{ij} \left|\left|\boldsymbol{R}_{ij}\boldsymbol{X} - \boldsymbol{D}\alpha_{ij}\right|\right|_2^2 + \lambda\left|\left|\boldsymbol{Y} - \boldsymbol{X}\right|\right|_2^2 + \sum_{ij} \mu_{ij} \left|\left|\boldsymbol{\alpha}_{ij}\right|\right|_0 \quad (3)$$

Here, $\mathbf{R}_{ij}$ is a matrix which indicates the ij-th patch of X. Therefore this first term controls the reconstruction error between the input noisy patch and the reconstructed patch. The second term ensures the relationship between the reconstructed image and the input image and the third term enforces sparsity to the solution. The parameter and $\mu_{ij}$ are weight coefficients that balance the sparsity and reconstruction rate. The goal of the parameter is to have the reconstruction error $\left|\left|\boldsymbol{Y} - \boldsymbol{X}\right|\right|$ below $C\sigma^2$. Where a is the standard variation of the noise and C is a constant.

The solution for the problem above is derived using OMP. Here $C\sigma^2$ is used as the parameter for the stopping rule of the OMP algorithm. Each target patches are reconstructed iteratively with the given dictionary **D** while the squared sum of the residual goes below the stopping parameter which is a function of the $C\sigma^2$. The patches derived from these routines are then aligned and overlapped parts are averaged linearly. Then the reconstructed image **X** will be the estimation of the original image **X** and this showed a great performances. Especially, the results using the dictionary from the noised image itself showed the better performance than the dictionary from natural images.

From the algorithm explained above, a simple question arises. Will it really give us the optimal solution? How the solution will change if the OMP algorithm had more or less iterations? Of course, less iteration will give less reconstructed image which may results in large difference between the reconstructed image the original image. In the same way, more iteration will give better reconstruction of the input image, which may lead to reconstruction of the noise.

Then what would be the optimal number of iterations? From the heuristic experiment, we found that the conventional algorithm which sum of square of the residual as the parameter for stopping rule does not give us the optimal number of the iterations. Now, the goal of our algorithm will be defining the better measure and parameters for the OMP algorithm's stopping rule.

To achieve this, we define the energy function to determine the noise level of the image. Therefore the noised image will have higher energy value than the original image. Here we assume that the energy function is linear so the condition stated below holds:

$$E(\mathbf{X} + \mathbf{N}) = E(\mathbf{X}) + E(\mathbf{N}) \quad (4)$$

If the condition above holds, the E(**N**) can be easily obtained from the training data by simply subtracting the energy value of the original image from the noised image. In this paper, the absolute sums of the gradients in x and y directions are used as the energy function. By comparing the energy values of the patches from training original images and the noised images, we can get the difference between the two sets. Using this value we estimate the energy value of the original image from the noised image. Of course, the energy function stated above, absolute sum of gradients, does not exactly holds the linearity condition above, but the condition fairly holds. In this way, the energy value of the original image can be estimated from the energy value of the noised image.

In our method, to denoise a noised image, we find the average energy value difference caused by the noise. Using the randomly generated patches, we add the same noise to the patches and compare the average noise differences between original patches and noised patches. Using the average energy value difference between noised image and original image, we reformulate (4) as follows:

$$\arg\min_{X,\alpha_{ij}} \sum_{ij} \left\|R_{ij}X - D\boldsymbol{\alpha}_{ij}\right\|_2^2 + \boldsymbol{\lambda}\left\|E(Y) - E(X)\right\|_2^2 + \sum_{ij} \boldsymbol{\mu}_{ij}\left\|\alpha_{ij}\right\|_0 \quad (6)$$

Here only the second term has been changed into the difference of the energy value from the squared sum of errors.

To solve the above equation, we again use the OMP. However, in the OMP algorithm, we use the difference of the energy values as the stopping rule instead of the sum of square error. In other words, we keep the iteration going until the energy function of the residual reaches to the average energy value difference. This will allow us to control the level of reconstruction so we avoid reconstruction the noise while estimating the original image.

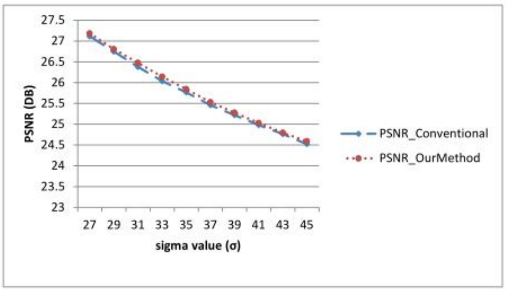


Figure 2. Comparison between the conventional method(blue circle, dashed line), and our method(red square, dotted line) using the global dictionary

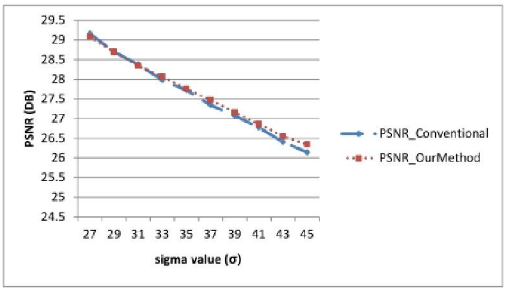


Figure 3. Comparison between the conventional method(blue circle, dashed line), and our method(red square, dotted line) using the adaptive dictionary

## Experimental Result

In this section we evaluate and compare the results from our algorithm with the conventional algorithm. The tests have been conducted on the Barbara image and the Boat with various noise levels. The exact same test has been conducted with [6], with the same dictionaries and noises for the comparison. Here the dictionary used was the dictionary from [6] which the size is 64 256. We used 8 8 image patches, which is the same size with the conventional algorithm. Table 1 and 2 compares the denoising results of our algorithm and the conventional algorithm. As the result shows, our method showed better performance in noise levels area for results using the global dictionary. For the results using the adaptive dictionary, our method show competitive result and gets better as the noise level and the complexity of the image increases.

**Table 1. Summary of the denoising PSNR result for Barbara image.**

| $\sigma/PSNR$ | Global Dictionary | | Adaptive Dictionary | |
|---|---|---|---|---|
| | Elad et al. | Our | Elad et al. | Our |
| 27/19.5 | 27.11 | **27.18** | **29.16** | 29.08 |
| 29/18.9 | 26.74 | **26.80** | **28.70** | 28.69 |
| 31/18.3 | 26.37 | **26.47** | 28.31 | **28.34** |
| 33/17.7 | 26.03 | **26.14** | 27.98 | **28.06** |
| 35/17.2 | 25.76 | **25.84** | 27.72 | **27.76** |
| 37/16.8 | 25.45 | **25.53** | 27.34 | **27.47** |
| 39/16.3 | 25.21 | **25.27** | 27.07 | **27.15** |
| 41/15.9 | 24.98 | **25.03** | 26.77 | **26.86** |
| 43/15.4 | 24.76 | **24.79** | 26.40 | **26.55** |
| 45/15.0 | 24.52 | **27.52** | 26.14 | **26.34** |

Table 1. Summary of the denoising PSNR result for Barbara image. Left column is the result using Elad et al. [6] and the right column is the result using our algorithm for various noise levels.

**Table 2. Summary of the denoising PSNR result for Boat image.**

| $\sigma/PSNR$ | Boat | |
|---|---|---|
| | Elad et al. | Our |
| 27/19.5 | 29.24 | **29.41** |
| 29/18.9 | **29.07** | 28.95 |
| 31/18.3 | 28.53 | **28.70** |
| 33/17.7 | **28.35** | 28.29 |
| 35/17.2 | **28.06** | 27.98 |
| 37/16.8 | 27.78 | **27.87** |
| 39/16.3 | 27.41 | **27.63** |
| 41/15.9 | 27.17 | **27.23** |
| 43/15.4 | 26.90 | **27.00** |
| 45/15.0 | 26.74 | **26.78** |

Table 2. Summary of the denoising PSNR result for Boat image. Left column is the result using Elad et al. [6] and the right column is the result using our algorithm for various noise levels.

## Conclusion

This paper has presented the algorithm developed from the conventional denoising method using sparse representation. It used the absolute of gradients as the energy function and used it as the stopping rule for the OMP algorithm. It showed slight but clearly better performances for the denoising using the global dictionary and competitive performances for the adaptive dictionary.
Yet, this algorithm has high potential to be improved by simply defining the better energy function for modeling the input noise. Also estimation could be developed by doing the regression about the energy values from the training data. This will clearly give us better estimation about the original image's energy value. Finally, this regression will allow us to model more complex noises such as noises from multimodal Gaussians.

## References


1. M. Aharon, M. Elad, and A. Bruckstein. K-svd: Design of dictionaries for sparse representation. Proceedings of SPARS, 5:9–12, 2005..
2. S. Boyd and L. Vandenberghe. Convex optimization. Cambridge university press, 2004.
3. P. Chatterjee and P. Milanfar. Clustering-based denoising with locally learned dictionaries. Image Processing, IEEE Transactions on, 18(7):1438–1451, 2009.
4. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. Image Processing, IEEE Transactions on, 16(8):2080–2095, 2007.
5. B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. The Annals of statistics, 32(2):407–499, 2004.
6. M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. Image Processing, IEEE Transactions on, 15(12):3736–3745, 2006.
7. J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1):1, 2010.
8. H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. Advances in neural information processing systems, 19:801, 2007.
9. Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on, pages 40–44. IEEE, 1993.
10. H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. Image Processing, IEEE Transaction on, 16(2):349-366, 2007